DOCUMENT RESUME

ED 167 580 95 TM 008 289

AUTHOR Cooley, William W.; Leinhardt, Gaea
TITLE The Instructional Dimensions Study: The Search for Effective Classroom Processes. Final Report.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Aug 78
NOTE 109p.

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Achievement Gains; Classroom Observation Techniques; *Compensatory Education Programs; *Effective Teaching; *Elementary School Curriculum; Grade 1; Grade 3; Individualized Instruction; Primary Education; Program Evaluation; Research Methodology; *Success Factors; Teaching Techniques; Time Factors (Learning)
IDENTIFIERS California Test of Basic Skills; *Instructional Dimensions Study

ABSTRACT

To evaluate classroom processes for teaching basic reading and mathematics, particularly individualized instruction in the context of compensatory education, a model of classroom processes was applied to disadvantaged children from 400 schools, grades one and three. The model specified that criterion performance is a function of initial student performance and four classroom processes: opportunity, motivators, instructional events, and structure. Criterion performance was measured by achievement gains on the Comprehensive Test of Basic Skills. A list of observable and measurable variables for each process was identified. Opportunity was the most powerful predictor of achievement gain. The most important findings for policy makers are the lack of clear evidence regarding the superiority of individualized instruction for compensatory education, and the insignificant relationship between extent of pull-out instruction and achievement. The strongest predictor of achievement gains, opportunity, refers to the degree of overlap between what was taught in a specific classroom and end-of-year achievement test content, as estimated by the teacher and by curriculum content analysts. The data imply that what is taught is more important than how. The appendix includes a rationale for the four processes and their respective variables, commonality analysis results, and correlations with gain for each predictor variable. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167580

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE INSTRUCTIONAL DIMENSIONS STUDY:

The Search for Effective Classroom Processes

Final Report

William W. Cooley and Gaea Leinhardt

Learning Research and Development Center
University of Pittsburgh

August 1978

TM008 289

INSTRUCTIONAL DIMENSIONS STUDY:
THE SEARCH FOR EFFECTIVE CLASSROOM PROCESSES

William W. Cooley

and

Gaea Leinhardt

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

PREFACE

This report presents a summary of the main results from the Instructional Dimensions Study (IDS) that are available at this time. Other reports have been and will continue to be produced by NIE and their other IDS contractors. In addition, we are currently conducting further analyses of those IDS data, and we hope those new results will further inform decisions regarding educational practices in compensatory education. Meanwhile, it seemed important to share these preliminary results at this time.

William W. Cooley and Gaea Leinhardt

August 1978

Instructional Dimensions Study: The Search for Effective Classroom Processes

In 1974, Congress directed the National Institute of Education (NIE) to perform a general examination of compensatory education programs. Included in that legislation was the specific directive to conduct a detailed study of the effectiveness of materials and procedures for meeting the educational needs of individual children. As part of this total effort NIE launched the Instructional Dimensions Study (IDS), a major survey of the impact of specific educational practices on student development. One of the IDS goals is to identify instructional procedures that are effective in teaching basic skills, particularly the use of individualized instruction in the context of compensatory education.

This report presents some of the results of the IDS Study. Specifically, it focuses on identifying effective classroom processes in regular classroom settings. We refer to "regular classroom settings" to distinguish them from supplemental classrooms, where the practice is to identify "disadvantaged children" and remove them from their regular classrooms for tutorial or small group work with a special teacher. It is our conviction that removing children from their regular classrooms does not adequately address the underlying problem of compensatory education, which is how to design educational

practices that are more effective in developing mathematical and reading literacy among children. If educational experiences in classrooms could be made more adaptive to the diverse needs of children in our schools, special programs for different categories of children (e.g., "disadvantaged") would not be necessary.

Therefore, it seemed appropriate to incorporate research that was designed to contribute to our "understanding of how current educational practice can extend its capabilities to respond to the diverse society it serves" (Glaser, 1977). The major objective of this aspect of the IDS is to identify classroom procedures that are particularly effective in teaching reading and mathematics to "disadvantaged" children in regular, primary grade classrooms, regardless of their family background or prior achievement.

Overview of Study

The design of IDS had several distinguishing features that relate to: the use of a model of classroom processes, direct measures of classroom processes, outcome measures, sampling procedures, and analysis procedures. Each of these features is briefly described below. The complete design is described in greater detail in Cooley & Leinhardt (1975b).

A model of classroom process. The Instructional Dimensions Study investigated classroom processes that are sensitive to differences in program designs and program implementation. The objective was to provide information on the instructional processes, especially individualizing ones, actually taking place in the selected classrooms, and on the influence of those processes on student learning. Because of Congressionally-specified interest, the study focused upon reading and mathematics instruction in the primary grades, using grades one and three to represent those grades.

The model for the study of classroom processes that IDS used is illustrated in Figure 1 (Cooley & Leinhardt, 1975a, Cooley & Lohnes, 1976). The figure indicates sets of variables that are assumed to be required for an explanation of the variation in student performance that occurs among classrooms after an extended period of instruction in those classrooms. It specifies that criterion performance is a function of initial student performance and of certain classroom processes that occur in the interval between the assessment of initial performance and the assessment of criterion performance, even though it does not specify the exact nature of those functional relationships. Classroom processes are represented by four constructs that are assumed to affect the criterion performance: Opportunity, Motivators, Instructional Events, and Structure. Aspects of individualization are described within the last two constructs.

Insert Figure 1 About Here

There are at least three general ways in which "model" is in current usage in this type of research: (1) general conceptual schemes which serve to identify the domains of relevant variables that should be "kept in mind" when considering the variables that might be included in exploratory studies of a particular phenomena; (2) formal structural models which describe specific causal relations that are assumed to exist among those major variables (e.g., path diagrams or structural equations); (3) statistical models which represent procedures for estimating the plausibility of the formal structural models. Figure 1 is a model only in the first sense.

Measures of classroom process. The model of classroom process was operationalized in the following fashion. Each construct was analyzed by reviewing extant literature, instrumentation and the results of our own previous research. A list of observable and measurable features for each construct was identified. The "best" measurement procedure(s) for each feature was selected. Best was defined as the least disruptive, most easily obtained, and most precise measurement. Measures within a construct were combined into variables. To analyze the data obtained in this way, measures are combined to form variables

which are in turn combined into sets of variables representing the four constructs.

It was assumed that classrooms differ in the manner and degree to which they are adaptive to individual differences in children. Classroom practices vary in how they provide for the variation in the knowledges and abilities with which children begin a school year, in how they respond to differences in student interests and motivations, in how they adapt to different rates of learning, and in whether or not they allow for a variety of educational goals. These variations are the result of: (a) differences in the instructional programs being used in classrooms, and (b) differences in the ways in which teachers implement any given program.

Outcome measures. The study utilized commercially available achievement tests in reading and mathematics. Available tests were assumed to be superior to any battery that could be constructed in the time that was available to mount the study, especially in terms of reliability and the elimination of racially biasing items, as well as in content validity. The test battery selected was the Comprehensive Test of Basic Skills of the California Test Bureau, level B used in grade 1 and level 1 used in grade 3. The rationale for selecting this particular battery is presented in Cooley and Leinhardt (1975b).

Sampling procedures. The primary objective of the IDS study was to identify effective mechanisms for improving educational programs. It was not intended to evaluate Title I or any

specific type of compensatory program. Therefore, the sampling scheme emphasized the need to achieve variance in classroom processes, rather than representativeness of current compensatory practices. This variation in processes was achieved by using descriptive information about the programs in use in a large number of school districts, and then selecting classrooms from those districts that would increase the likelihood of process variation. Other programmatic considerations in sampling included the stability of the instructional program (programs new to the school were excluded) and the setting variable (the nature and degree of pull-out instruction taking place).

A detailed description of the sampling procedures is available in Kirschner (1977). The multistage sampling process resulted in 400 classrooms in 100 different schools, from 14 school districts, in 5 states. In addition to a wide representation of curricular and instructional approaches, the sample was diverse in geographic and demographic features such as district size, location, and degree of poverty.

Analysis procedures. The analytic procedures used were based on the recognition that classroom practices occur in various combinations in the field that are not controlled by researchers. What is needed at this stage of our understanding of effective classroom practices is a technique for determining the relative importance of different practices in explaining the differences in student achievement. One specific technique that accomplishes this is commonality analysis, a variation of

regression analysis that allows one to distinguish between unique effects and effects that are common to two or more of the predictor variables, where the predictors are classroom process dimensions and children's initial abilities. This technique provides the basis for inferences about the relative usefulness of different classroom processes in predicting student achievement, and the extent to which useful classroom process predictors reflect methods of individualization.

Another important feature of the data analysis scheme in this report is that the classroom is used as the unit of analysis. Thus, measures on individual students are aggregated at the classroom level. For example, classroom means and standard deviations are used to describe the class distribution on pretest scores.

Data Collection Techniques

Three data collection techniques were used to assess the process variables: (a) interviews of the teachers, (b) analysis of the curricula by curriculum experts, and (c) videotapes of classroom activities. Two of the three techniques (interviews and taping) had been developed and utilized prior to the design of IDS, during extensive research on the impact of classroom processes on achievement (Leinhardt, 1977a) in the context of Follow Through, a compensatory education program.

Teacher interviews were used to determine specific classroom practices. Interviews were given twice during the study. The information from the interview was supplemented by records of student assignments and pupil curriculum records (PCR) collected at later times. Curriculum analyses were made based on the PCR's which informed the analysts of the specific instructional material to which the students were exposed. The analysts then made judgements about the structure and quality of the curricular materials.

Videotaping provided unique information about classroom practices, particularly about the quality of the Instructional Events. Taping required fewer highly trained observers than in-class observation, eliminated the possibility of confounding observers with sites, and provided a permanent record of activities that made it possible to monitor coding accuracy, recode ambiguous results, and reanalyze data at a later time using different coding schemes. Teachers were taped for approximately four hours of active instruction. Details of the collection and coding for all instruments is available in Kirschner (1977).

There were distinct advantages to using this three-pronged procedure. Information was gathered in such a way as to permit careful, reliable analysis in a different location from the classroom. Also, permanent records were created that will be available for reanalysis using different analytical approaches and asking different questions of the data. Finally, the most

efficient (economic and precision preserving) procedure was used to gather each piece of information.

Overview of Measures

An overview of the measures used, the instruments used to collect them, and the manner in which they were combined into variables is presented in Figures 2 through 5. A more detailed description of the classroom process measures that were used in IDS, with a rationale for how they were combined into variables (i.e., composites of sets of measures), and how the variables were clustered into each of the four constructs, is provided in the Appendix.

In Figures 2 through 5, the first column identifies the variable and the second column lists the set of measures that form that variable. A minus sign following a measure indicates that the measure is added negatively when forming the variable. The last set of columns identifies the data collection source of that particular measure.

Figure 2 lists the measures for the Opportunity construct. A classroom would score higher on the time variable if it tended to have: lower enrollment, high attendance, more minutes scheduled for reading or math, fewer transfers in or out, more instruction time per year, and a higher percentage of students on-task. A classroom's score on the curriculum overlap variable is a function of both the teacher and curiculum analysts

estimate of the degree of overlap between what was taught and what was in the end-of-year achievement test.

Insert Figure 2 About Here

Figure 3 lists the measures for the Motivators construct. A classroom scored higher on the motivating curriculum variable if it tended to have: interesting curricular materials, several modes of instruction, and rapid correction of student work. A classroom scored higher on the interpersonal motivators variable if it tended to use: peer tutoring, games and contests, student self-evaluation, student self-management with respect to seating and working, and more teacher praise for specific and nonspecific behaviors.

Insert Figure 3 About Here

Figure 4 lists the measures for the Instructional Events construct. A classroom tended to score high on the inappropriate management strategies variable if the teacher made a large number of management statements to groups or individuals. A teacher scored higher on the teaching to individuals variable if she or

he: made more statements or questions to individuals, elicited more individual child responses, had a higher ratio of cognitive to management statements, contacted a large number of individual students, or tutored individuals. A teacher scored high on the teaching to small groups (or the whole class) of students to the extent that he or she made contacts of the same type, but to small groups (or the whole class) of students, and to the extent that he or she discussed tangential information (or minimized lecture time). The teacher scored high on the quality of instruction if he or she: had fewer silences, sought active student responses, referred to earlier curricular material, focused the child's attention, solicited the child's opinion on correctness of the child's response, referred the child to previous successes, tutored effectively, managed the class effectively, gave prompt and accurate feedback to the student, and interacted with individuals.

Insert Figure 4 About Here

Figure 5 lists the measures for the Structure construct. A classroom scored higher on the placement of students variable if it tended to do more of the following: had placement or pretests available from the curriculum; used standardized, curricular-built, or teacher-built tests; or used a student work sample to assess placement. A classroom scored higher on the

systematic individualized assessment of mastery of curricular content variable if it had or did more of the following: had posttests available from the curriculum; utilized curricular methods or tests; or utilized teacher tests, in-class assignments, completion of a work unit, observation of a work sample, or a student interview to assess mastery. (The purpose of this variable was to determine if any systematic assessment procedures were used by the teacher.) A classroom scored higher on the individualized matching practices variable if it had: fewer days between tests; a smaller range of days between tests; a higher proportion of students doing well on tests; more frequent testing; and more frequent use of tests to assess mastery, or diagnose student need, and/or make prescriptions. (The purpose of this variable was to determine what matching techniques were utilized, and how, by the teacher--including, but not only, mastery tests. If a teacher assessed mastery only by testing and used tests solely to assess mastery, then these two measures in two different variables would the same.) A classroom scored higher on the group-and-assign-students-individually variable if it had: a higher percentage of unique assignments; more assignments made individually; relatively little time spent in whole class instruction or in groups greater than 11; more time spent in small groups of less than four, or in one-to-one instruction; fewer students in groups; and frequent regrouping. A classroom scored higher on the individual needs variable if, when the student had difficulty with the curriculum: the curriculum

provided more additional practice or rules for repetition of material; the teacher provided for repetition, remedial material or tutoring were more frequently used, or there was more opportunity to skip ahead and return later; and reading materials had a more appropriate readability level. A classroom scored higher on the individualized sequence of instruction variable to the degree that the instruction was: clearly sequenced, the clear sequence was followed, sequences varied within the classroom, weak sequences were improved by teacher-made supplemental materials, sequences were prerequisitely ordered, sequences were altered to meet student performances, instruction was self-paced and curricular rules guided pacing, or curricular content was systematically reviewed.

Insert Figure 5 About Here

In addition to the variables that relate to the constructs of Figure 1, another variable called setting was included in the analyses reported here. The setting variable is a function of two measures, the number of children pulled-out of a classroom for special instruction in reading or mathematics, and the amount of time devoted to such pull-out instruction. The variable was included to determine whether or not the degree of pull-out instruction made a difference in achievement for the class as a whole.

Reducing and Analyzing Data

This section summarizes the procedures that were used for selecting the classrooms actually used in the analyses, reducing classroom process measures to a small number of variables, and analyzing the resulting variables to reveal the usefulness of various instructional dimensions in predicting reading and mathematics achievement.

In order to create a usable data set, classrooms with large amounts of missing data, or consistently aberrant values, had to be deleted from the sample. Each of the four sets (i.e., first and third grade reading and mathematics) of classrooms was "swept," and those classrooms that were missing all of the data from one or more data sources (videotapes, interviews, analyses, etc.) were deleted for all analyses in that set. The most common sources of data that were missing were videotapes and curriculum analyses. Classrooms were retained in the sample if only one or two measures from within a data source were missing.

Reducing Data

Describing and analyzing the voluminous information collected in IDS required that the classroom data be systematically reduced to a manageable number of process dimensions. For example, it would be conceptually impossible and statistically undesirable to analyze over 100 measures separately. Fortunately, the measures described earlier were constructed with a solution to this problem in mind. The basic

approach was to reduce the measures along the lines of the six constructs of the classroom process model illustrated in Figure 1. The steps involved in data reduction were: elimination of unusable measures, preliminary correlation and partial correlation analysis of measures with pretest and posttest, inspection and reflection of measures, development of standard scores with unit variance, and combination of measures to form variables. By combining the data in a manner that preserved the initial "meaning" of the measures, the final results were more interpretable than if they were combined simply on the basis of interrelations among the measures.

Elimination of measures. In any large-scale data collection effort, some of the information will be unusable for one reason or another. The first step eliminated such measures. Means, standard deviations, and skewness were examined, and measures with zero or near zero variance were deleted, specifically those measures that had 85% or more of the data at a single value. Measures with unusual skewness were examined for outliers to check for obvious errors. We also computed frequency distributions and scanned them for "unusual" values. Obviously, not all unusual values are incorrect, but simple clerical errors were detected and corrected in this way. A list of deleted measures and the reasons for the deletions are presented in Tables 3 through 6.

Correlation and partial correlation. The next step in data reduction was to inspect the correlations among the measures for each of the variables. Correlations were computed on the measures along with data from initial and end-of-year student performance aggregated as classroom means. These correlations were inspected, particularly to identify unusually high relationships between initial abilities and process measures which might unnecessarily confound process and initial abilities.

The signs of relationships between measures and outcome residuals also were inspected. This was accomplished by partial correlation, by removing initial ability from process and outcome. The objective was to identify measures that did not appear to conform to theory with respect to their effects on achievement. (This procedure is a conservative form of criterion scaling.)

Combination of measures. The procedure used for combining measures into variables was to add the measures for each variable after they had been adjusted to unit variance, with separate scalings for math process and reading process at each grade. The obtained value for the combined measures was then divided by the number of measures in the set to create a mean, thus permitting adjustments to be made if one of the measures in a variable was missing. The complete procedure reduced the classroom data from 95 measures to a manageable number of 15 variables. These process variables were then combined with initial abilities, a setting variable and outcomes. Four separate regression analyses

were conducted: grade 1 reading, grade 3 reading, grade 1 math, and grade 3 math.

Analyzing Data

The original RFP for IDS requested a plan "for arguing causal relationships between program and outcome variables" in a nonmanipulatory survey of existing classroom practices, despite the difficulty of such arguments under the canons of statistics (as reviewed by Cronbach & Furby, 1970). Since it is well known that correlation cannot prove causality, what was required were analyses that created the strongest valid presumption of causality. Although there are analyses of correlations that permit valid and stronger presumptions of causality under ideal research conditions (e.g., path analysis), the actual conditions for the IDS called for a conservative method of analysis. On the technical side, this was because the measurements employed possessed only partial validities for the constructs they represented, and had variable and unknowable degrees of unreliability. On the theoretical side, the available causal models for explaining differences in student achievement are still too primative to avoid specification error, which is a great threat to causal interpretations of regression coefficients. (See Lohnes and Cooley, 1978, for a more extensive discussion of this problem.) On the policy side, conservatism is desirable because there is probably less hazard to the nation when research underestimates (for Congress) the impact of particular arrangements for education, than when research claims

unreplicable effects. Commonality analysis provides an appropriately conservative approach to analyzing the results of the study.

Commonality analysis is a comparison of statistical models for the data. When it creates a presumption of causal efficacy for process variables, it does so by reducing the probabilities for alternative explanations of the observed educational development of pupils. The basic fact about survey research data is that the possible causes of outcomes, as described by alternative models, are confounded to some extent. Rather than attempt hazardous discrimination by algebraic tricks, commonality analysis presents the confounded portions of effects on outcomes separately (called "commonalities"). These confounded portions are not included in estimating the irreducible effects of the separate possible causes (called their "uniquenesses"). The conservativeness of the method resides in the use of minimum rather than maximum estimates of separate effects. The advantage to the method rests with the visibility of the confounded effects of variables of sets of variables. Other conservative approaches tend to mask them.

While the external validity of a survey study is dependent on the representativeness of the sample, the internal validity is dependent on the qualities of the models compared. Causal arguments will be most convincing in the presence of complete and technically competent, theoretical models of classroom learning. The omission of probable causes that are correlated with the

hypothesized causes is devastating. Unconvincing operationalizations of constructs or serious unreliabilities are debilitating. Exhaustiveness and redundancy are the rules to follow in conducting observational studies. Parsimony and organization are achieved by combining detailed measures prior to commonality analysis. That is, the classroom process variables are scaled as linear functions of a set of related measures.

The model that has guided these analyses postulates that a particular set of process variables that describes ways and degrees to which instruction is individualized, is essential to the explanation of the pupil learning that occurs in reading and mathematics during a school year. The model holds that the four sets of process variables combines with the initial abilities to generate pupil learning. The alternative models hold that: (a) initial abilities are sufficient explanation of pupil learning, and (b) abilities together with process variables not related to individualization are sufficient explanation of learning. Since these are simpler explanations, their truth probabilities can only be reduced by showing appreciably greater power for the complex model involving the addition of information about individualization.

Results

The analyses that follow are based upon four sets of classrooms: grade 1 reading, grade 1 math, grade 3 reading, and grade 3 math. Table 1 displays the overlap between the two sets of classrooms at each of the two grade levels. The classrooms in the analyses are about one-half of the total IDS sample, and are restricted to those classrooms for which reasonably complete data were available. Incomplete data occurred primarily because budget constraints prevented the videotaping of all classrooms in the study. The basis for deciding which classrooms were videotaped is described in Kirschner (1977).

Insert Table 1 About Here

The top half of Table 1 describes the grade 1 sample. Of the 200 first grade classrooms in the IDS study, 104 were included in the reading study. Of these, 51 classrooms were also among the 86 in the first grade math analysis. The bottom half of Table 1 describes the grade 3 sample in a similar fashion.

Table 2 displays the correlations for the total scores on the Comprehensive Test of Basic Skills (CTBS) for fall pretest, spring criterion performance (posttest), and the difference between pretest and posttest (gain). Note the high degree of

correlation between pretest and posttest for grade 3. It suggests that, in spite of large differences in classroom processes operating in third grade, classrooms tend to have the same general rank-ordering on this achievement test at the end of the year as they had at the beginning of the year.

Insert Table 2 About Here

The pretest-posttest correlations were not as high in first grade. Thus in first grade it is more likely that one can identify classroom process differences that are making a difference in how much improvement is being achieved in reading and mathematics. It is also possible that initial performance was not measured as reliably in first grade classrooms, resulting in lower correlations between pretest and posttest.

A total of 95 measures was retained to describe the differing processes used in the classrooms included in the study. (A rationale and brief description of each measure is presented in the Appendix.) As described earlier, the process measures were combined to form variables, which in turn were grouped into the four constructs of the model. Tables 3 through 6 show the means and standard deviations by construct of the measures included across all four samples; each table also lists the measures that were deleted and lists the reasons for deletion.

The means and standard deviations for a measure within the Opportunity construct in Table 3 tend to be similar in value across the grades and subjects. There are exceptions to this general trend, however, which are important to consider. Almost twice as much time is allocated to reading instruction as is allocated to mathematics instruction. In addition, the teachers' estimates of the overlap between curriculum and criterion were all quite similar except for grade 1 math, which was much lower than the others. Thus, interpreting the effect of overlap for first grade math was very different than interpreting the effect for third grade math. What the data suggest is that the criterion test battery chosen for first grade math tended to have less than one-third of its items taught. A classroom with high overlap in first grade math still had less than 50% of the criterion material covered.

Insert Table 3 About Here

The means and standard deviations for the Motivators construct are presented in Table 4. In general, the means and standard deviations are consistent across the grades. There are slightly fewer modes of instruction used in math than in reading, somewhat less praise given in third grade than in first, and slightly less self-evaluation in first grade than in third.

Insert Table 4 About Here

The means and standard deviations for the Instructional Events construct are presented in Table 5. In general, the means were similar across all four sets of classrooms. Some features of the data are interesting. More contacts of all types (except small group) are made in math instruction than in reading. Also, in mathematics more contacts were directed to large groups than in reading (by a margin of 3 to 1), while in reading contacts were directed to small groups and individuals. For the measure small group child responses, the frequency was considerably higher in first grade reading than in the other sets. This may be due in part to the influence of the Dystar reading program at that grade, as it is one of the few programs that specifies group responses, and was widely used in the first grade reading sample.

Insert Table 5 About Here

The means and standard deviations for the Structure construct are shown in Table 6. In general, math classes had more whole class instruction, larger groups for grouped instruction, and less individualized instruction, than reading

classes. A disturbing picture of classroom practices is provided by the measure of unique assignments. The measure is an estimate of the percentage of assignments within a class that were unique for each student in that classroom. The average values were very low (12-21), indicating a meager degree of individualization in even the most individualized classes. In a different series of studies of individualized programs (Leinhardt, 1976, 1977a, 1977b), uniqueness was approximately 80% (S.D. = 10). In this sample, only 23 classes achieved this degree of individualized assignments.

Insert Table 6 About Here

Outcome Variance Explained by Full Model

Using spring achievement as the dependent variable, multiple correlations were run for each of the four sets of classrooms, using fall pretest and all classroom process variables as the predictors. Table 7 reports the number of classrooms in each analysis, the multiple correlation for predicting class mean achievement, and the proportion of the variance in classroom means that is explained by the 18 predictors (two input variables, 15 process variables, and 1 setting variable). As shown in that column, about one-half of the end-of-year variance in mean achievement was explained for first grade, and three-fourths was explained for third grade.

The last column reports the R^2 values after correcting for shrinkage. It adjusts for the fact that there will be some overfitting when 18 predictors are used with samples of the size indicated. The largest adjustment was for grade 1 mathematics because of the smaller sample size for that group.

The importance of this table is that it indicates how close we are getting to an explanation of the differences among classrooms that affect end-of-year performance on a standardized achievement test. It is clear that we are much closer in grade 3 than in grade 1, but this is due to the higher correlations between pretest and posttest for grade 3, as was seen in Table 2. The next task, then, is to sort out which sets of predictors are most useful in this explaination of student achievement in reading and mathematics. Commonality analysis was used for this purpose.

Insert Table 7 About Here

In the commonality analyses reported here, achievement gain (posttest minus pretest) was used as the dependent variable. Analyses were also run using posttest as the dependent variable, but the results were reported using gain for the following reasons. The tangible presence of actual student gain, the

reality of variation in gain across classrooms, the lack of serious confounding between gain and pretest help to focus the question of how education makes a difference. This helps to emphasize that while much of the variation in end-of-year achievement is explainable by initial performance, this is not the same as saying that school did not affect student growth. Further the relative importance of the process variables was the same whether student gain or posttest performance was used as the dependent variable. In short, we viewed gain as a more interpretable and useful dependent measure.

Consideration was also given to using the unbiased estimate of true mean gain scores, rather than obtained mean gain scores. However, as Davis (1964) points out, obtained gain is essentially equal to estimated true gain for grouped data. For these and other reasons it was decided to use raw mean gain in the commonality analyses which follow.

Results of Commonality Analysis

The major trends in this vast collection of data can best be seen in Tables 8 through 11, which summarize the results of the commonality analysis of the four samples of classrooms. The first column in each table, the uniqueness, provides an indication of the usefulness of that particular set of predictors in predicting gain. A uniqueness is the increase in the R^2 that occurred when that set of predictors associated with a particular construct was added last, where the square of the multiple

correlation coefficient is the proportion of variance in achievement gain that is explained by all six sets of predictors.

Insert Table 8 About Here

The variables included in the Opportunity, Motivators, Instructional Events and Structure constructs have been previously described, but the pretest and setting constructs require additional explanation. The pretest construct consists of two variables. One is the classroom means on the fall pretest, the other is the within classroom standard deviation. The latter was included to establish whether or not the amount of gain achieved in a classroom was in part a function of the degree of heterogeneity in the abilities of the children in that classroom.

The setting construct is represented by a single variable. It is the percent of total instruction that was provided outside that classroom. It reflects both the number of students receiving supplemental ("pull out") instruction as well as the amount of time spent in that other setting. It was included to control for the possible effect of supplemental instruction on the whole classroom.

The commonalities indicate the degree of confounding that occurred because of correlations among the predictors. For example, in Table 8, Opportunity and Structure (sets 2 and 5) had a common effect of .03 for grade 1 reading. This was because the Opportunity variables had a slight tendency to be correlated with the Structure variables, and there was no way of sorting out this ambiguity statistically without making an arbitrary decision in favor of one predictor set over the other. These commonalities can be thought of as indicating the potential usefulness of the predictors. That potential could be realized if it were possible to obtain lower correlations among the predictors through sampling strategies or intervention in subsequent studies, or by reconsidering the procedures for forming variables from measures.

The last column in the tables reports the probabilities for the uniquenesses being equal to zero. Although these are not independent tests of inferences, they do incorporate information about the number of predictors in the construct, and thus give an indication of how much confidence we can have in that construct's usefulness in predicting gain.

At the bottom of each table is the squared multiple correlation coefficient, which is the variance in gain that was explained by all six constructs together. Also shown is the probability that the R^2 was greater than zero. The P values help to determine whether there might have been some overfitting in estimating these multiple regressions. One can obtain large R^2's

simply by having a large number of predictors relative to sample size.

Also at the bottom of Tables 8 to 11 is an estimate of total unique contribution of the four process constructs. These are helpful in getting an indication of how useful the combined set of classroom process measures are in predicting gain, independent of pretest and setting. These process uniquenesses include the uniqueness of each process construct as well as commonalities among the four constructs. They represent the amount that the total R^2 would be reduced if the process variables were not included in the regression.

The first grade results were similar in that, for both reading and math, the pretest and the Opportunity variables were the most useful predictors of gain in achievement. In first grade, Opportunity was by far the best predictor of reading gain; whereas in math, the pretest and Opportunity had identical uniquenesses.

In Table 9 the usefulness of commonality results can be seen. Here, the Motivators construct shows promise because of the combined effect it has with pretest. This suggests that Motivators may indeed be important in first grade math, but we cannot be certain because the motivator variables were related to pretest scores.

Insert Table 9 About Here

Third grade reading (Table 10) was the sample which had the highest correlation between pretest and posttest ($r = .86$). Thus, with almost three-fourths of the variance in posttest associated with pretest, there was less reliable variance left for other predictors to explain. Even so the usefulness of pretest and opportunity in predicting gain are comparable to grade 1 mathematics. One way in which the third grade reading results differ is in the importance of setting. We are not sure why this is so.

Insert Table 10 About Here

Third grade math (Table 11) also shows the potency of the Opportunity construct. In fact, this was the only construct that had a significant unique ss. In the other three samples at least one other construct did.

Insert Table 11 About Here

In summary, the major generalization with respect to classroom processes must be that the most useful construct in explaining achievement gain is the Opportunity that the children had to learn the skills assessed in the achievement test. Motivators showed some promise for grade 1 math and grade 3 reading, and Structure showed promise for third grade math.

Tables 12 through 15 provide additional ways of looking at the relationships between the predictors and achievement gains for the four sets. The first column provides the standardized partial regression coefficients (b) for predicting gain from the 18 predictors. The second column reports the zero-order correlations between each of the 18 predictors and gain. Here it is quite clear that the curriculum overlap measure within Opportunity was the important predictor of gain. It is also interesting to note that in three of the four sets the pretest means were negatively related to gain. This needs to be studied further, as it may simply be a regression to the mean artifact that is a result of the unreliability of the estimates of initial ability for first graders.

Insert Tables 12, 13, 14, and 15 About Here

An examination of the zero order correlations reported in Tables 12 through 15, reveals some interesting patterns. For example, consider the highest correlations between a variable and the gain criterion. First, for all four sets, curriculum opportunity is the most highly correlated with achievement gain. The next most highly correlated variable is the Structure-matching variable for mathematics, and Opportunity-time for reading. In terms of the regression coefficients, the importance of curriculum Opportunity is once again the most clear and consistent result.

Results at the Measure Level

Inspite of the lack of strong relationships between three of the four process constructs and student gain, it may be worthwhile to summarize the important relationships between individual measures and gain. Although we feel that the simple sum of the standardized measures is at present the best way to combine measures to form variables, it is not without its drawbacks. If the set of measures for a variable have relatively high positive intercorrelations among themselves, then the linear combination of them will enhance the reliability of the resulting variable. If, on the other hand, the relationships are low or even negative (as is the case when teachers use one of two possible practices, but not both) then the combination process tends to lose information. This can be demonstrated by the observation that in no case except for overlap was the correlation between the composite variable and criterion as high

as the best measure in the set. The following paragraphs summarize important relationships at the measure level that are not revealed by the previous inspection of variables.

Time. The most surprising finding is that the amount of time scheduled for a specific subject matter bears so little relation to gain or raw achievement. Clearly, utilization of time (an estimate of which is provided by % on task) is more important than allocation, although allocation is generally presumed to be the upper bound of utilization. (Note: a trial attempt to combine the measures multiplicatively, number of days times number of minutes times percent on tasks did not improve the situation.) Another unexpected finding is the lack of a consistent relationship between class size and achievement. Large classes may make the environment less pleasing and work conditions less tolerable, but they do not systematically detract from learning within bounds of our class size.

Assessment of mastery and matching. Assessment of exactly how much students have mastered seems to be fairly important for student gain especially at the 3rd grade. What is interesting is that assessment of mastery by a formal in-class assignment or a curriculum provided method are as effective as testing. It also is important, however, to test students frequently on material they have been taught even though other techniques may be used to assess mastery of material. One possible explanation is that frequent testing on learned material gives a student the opportunity to practice taking tests. This practice consists of

both rehearsals of test taking skills, but also of opportunities to cope with the anxiety and discipline associated with even short quizzes.

Management information. With the exception of first grade math, the measures of amount of management information given to students is negatively related to student achievement. Obviously, students need some management information but too much of it detracts from cognitive instruction.

Quality of instruction. The specific techniques that appear to be most relevant for quality of instruction are: focusing a child's attention on the task, referring to earlier curriculum materials, referring to earlier performance, and effective classroom management.

Implications

In this section we turn to the implications of the study for educational policy, instructional practice, and evaluative research. The implications for educational policy focus on some of the broad national decisions for which this study might have relevance. The implications for instructional practice focus on the concrete classroom activities that seem to affect student growth. Finally, the implications for evaluative research deal with the unresolved issues that need further work and the ways in which this approach have been particularly fruitful.

Policy Implications

Probably the most important finding of the Instructional Dimensions Study is the fact that there was no clear evidence regarding the superiority of individualized instruction for compensatory education. That is, the simpler model of pretest plus process variables not related to individualization is sufficient to explain student growth. Although, some of the measures that reflect various aspects of individualization did show some effect in one or more of the samples studied, there was no consistent pattern across the four samples. These studies seem to be consistent with other attempts at establishing the effectiveness of individualized instruction. The policy implication is that it seems premature for Congress to identify individualized instruction as the approach to be used in compensatory education. This was suggested when the ESEA

reauthorization was under consideration in 1974 and that suggestion prompted Congress to request a special study of the effectiveness of individualized instruction.

These results should not be interpreted as indicating that a dead end has been reached in attempting to adapt classroom instruction to individual differences in children. There are two major problems which this study has revealed. First, the measures of individualization are still not sufficiently developed to provide the reliability needed for detecting possible program effects. Secondly, although the range of individualizing processes that were included in the sample are probably representative of those currently available in the field, they are still far from the theoretical ideal with respect to what is possible in individualization. For example, the measure of percent of unique assignments revealed that these percentages are quite low. In more controlled implementation of one form of individualized instruction, Leinhardt (1977a) found percent uniqueness to be as high as 85 percent.

Another finding with policy implications is that there was little relationship between the extent of pull-out instruction that was taking place in classrooms and the achievement of children in those classrooms. The setting variable, which reflected both the number of students receiving pull-out instruction and the amount of instructional time that was involved in such instruction, was related to student gain only in the case of third grade reading. Thus, neither the students

pulled out nor the classmates who may have benefited from reduced class size during the pull-out period seem to have shown much improvement over classrooms in which little or no pull out was occurring. Or possibly, one group's benefits matched the other group's losses.

Finally, the most pronounced trend in these data, the importance of opportunity to learn, suggests that the most useful thing to do for children with underdeveloped reading and mathematics skills in the primary grades is to provide more direct instruction in these areas. Of course, given a finite school day, more of something means less of something else. This suggests that more consideration needs to be given to the planning of the day for individual children, such as exploring how more time might be allocated to higher priority objectives. It seems clear that the more important consideration is what gets taught, rather than how it's taught. This fact alone might be very important in improving the educational experiences of children having difficulty learning the basic skills in the primary grades.

The implications that have emerged would lead us to recommend a more intensive re-examination of the content of school curricula, and the content of examinations. If the success of Compensatory Programs is to be judged in terms of their impact on what students know, which is estimated by how students perform on tests, then it behoves Congress, State Departments of Education, and school districts to devote

considerable effort to the definition of what children should be learning and what should be in tests designed to assess learning. Issues surrounding differences in teaching method (open classrooms, personalized instruction, deductive vs. inductive) seem to have distracted us from the more central concern, what is important to teach. In sum, both the content of instruction and the content of tests should reflect those things that are valued by the society. The research community can contribute to resolving what is and ought to be valued by society and how that can be translated into curricula and measurement. This is not to imply the need for a single, national curriculum. What is important is that what is taught is consistent with what is tested, and that both are justified through a rational process.

Instructional Implications

The results of this study have some definite implications for instructional practice. As with other recent studies of this type (MacDonald & Elias, 1976), we can see from the results that what goes on in a classroom does have definite impact on what students learn during the course of an academic year. At least one-fourth of the variation in achievement gain is due to differences in classroom processes. What is less clear is precisely which classroom practices make what kind of a difference with whom. We have the beginnings of an explanation of this and we will attempt to sketch it out below.

The importance of curriculum content. Instructional techniques, both curricular and interpersonal, do not compensate for missing content. Children perform on a test much better if they have been exposed to both the form of the test and the content covered by the test. This point may seem trivial but it is in fact quite important. Both what is taught and tested represents a sampling from domains. Curriculum can clearly emphasize different content than does the test. An assumption is made about the ability of students to generalize and the ability of tests to accurately sample both what has been directly taught and what the student should be able to generalize to. The results of IDS show that students are more likely to answer correctly if they have been directly taught the material covered by a test and if they have been frequently exposed to the test format.

What the data seem to indicate is that there are many different ways of teaching and that no one way--individualized or grouped--is superior. When certain ends are met such as assessing student mastery of material or responding to individual student needs, the question isn't how it's done, but that it is done in some fashion. No one technique of instruction is clearly associated with disastrous outcomes or successful ones. Another way of putting this is that there are compensating educational practices with respect to the association of techniques and student differences in ability.

To manage or to teach? The results support the idea that the emphasis in instruction should be on the cognitive rather than on the managerial. When teachers are forced either by overly complex programs, poor class discipline, or poor general organization to focus on classroom management, they do so at the expense of direct instruction. This is supported by the following: consistent negative relationships between management contacts and cognitive ones; generally negative relationships between management and gain; generally negative relationships between management statements and the quality of instruction.

This finding relates to the fact that it is more important what is taught than how it is taught. An elaborate, innovative curriculum is unlikely to be effective if it requires sustained (as opposed to start up) managerial time and effort on the part of the teacher because such time is obtained at the expense of instruction. Given the indirect evidence in this study that most classrooms utilize several curricula simultaneously, it is important to recommend a usable system for incorporating them.

Implications for Program Evaluation

Of the four process constructs assessed in IDS, Opportunity is the strongest, most consistent predictor of achievement gains. Within Opportunity, the variable which assesses the degree to which the children in the classroom had the opportunity to learn what was in the end-of-year achievement test stands out. That variable is a composite of two measures. One measure is based

upon the teachers' estimate of what was covered during the past year and the other measure is the result of a content analysis of the curriculum materials used in the classroom during the year.

The predictive validity of this curriculum overlap variable stands out no matter how the data are analyzed, and in terms of mean- gain- variance explained, it goes up to 20% for first grade reading. So once again we are faced with this finding: children tend to learn what they have the opportunity to learn, they tend not to learn what they do not have the opportunity to learn. It seems appropriate to come to grips with the implications of this finding for program evaluation.

One implication of the importance of Opportunity is that evaluation studies which do not include information on how time is allocated or on the degree to which the curriculum included what the achievement test measured are suffering from specification error. That is, there is the danger of attributing instructional effectiveness to specific programs or ways of teaching when it is really a matter of the curriculum content being a good fit to the particular achievement test that happened to be selected.

Another implication of the potency of Opportunity is the need to find ways of reducing the arbitrary nature of the selection of measures of school outcomes used in program evaluations. Including overlap information may reduce the likelihood of specification error, but it is still not a

completely satisfactory solution. Work on establishng the transfer value of the outcome measures used is needed.

The Instructional Dimensions Study illustrates an approach to program evaluation that considers educational treatments as a multi-dimensional domain. But this experience with IDS also points out that we still have a good deal of work to do in developing adequate measures of the major dimensions of classroom environments. The problem is indeed complicated. If one identifies over one hundred specific ways in which teachers and curricula can vary, it is necessary to have ways of combining these measures into composites prior to fitting to criterion. If this is not done, one either over-fits the data or requires very large samples of classrooms. It is unlikely that the research community can attract the resources necessary for samples larger than IDS.

A variety of ways of combining the observed measures into variables has been tried and so far none of them has been completely satisfactory. Factoring has tended to produce meaningless dimensions, although they do have the nice property of being orthogonal. When interpretable dimensions are created, orthogonality and validity tends to be lost. One or more of the specific measures in a composite generally had larger predictive validities than did the composite itself. The scaling of the dimensions of classroom environments seems to require methods that are different from those developed for assessing individual differences. For example, sometimes two measures of the same

classroom characteristic are negatively correlated because they represent different ways in which a teacher might accomplish a particular task such as placement testing. It might be done in one of two ways, but not both. In another set of measures, which indicates the teacher's motivating strategies, a variety of options are available, and whether or not a teacher uses one of them is independent of whether other options in that set are used. Another set of measures may have more familiar psychometric properities, in that the measures that define a specific variable have modest positive correlations, and their linear composite has a larger predictive validity than any single measure of the set. (Measures of curriculum overlap had that nice property.) We have been trying to apply a single scaling method across all sets of measures but it now appears that each set of measures may have to be scaled independently. But however it's done, it is one of the most important challenges facing program evaluation today.

References

Cooley, W. W., & Leinhardt, G. Design for the Individualized Instruction Study: A study of the effectiveness of individualized instruction in the teaching of reading and mathematics in compensatory education programs. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975. (a)

Cooley, W. W., & Leinhardt, G. The application of a model for investigating classroom processes (LRDC Publication 1975/24). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975. (b)

Cooley, W. W., & Lohnes, P. R. Evaluation research in education. New York: Irvington Publishers, 1976.

Cronbach, L. J., & Furby, L. How we should measure "change"--or should we? Psychological Bulletin, 1970, 74(1), 68-80.

Davis, F. B. Educational measurements and their interpretation. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1964.

Glaser, R. Adaptive education: Individual diversity and learning. New York: Holt, Rinehardt and Winston, 1977.

Kirschner Associates, Inc. Instructional Dimensions Study: Final Report. Washington, D.C., 1977.

Leinhardt, G. Applying a classroom model to instructional evaluation. Curriculum Inquiry, in press.

Leinhardt, G. Observation as a tool for the evaluation of implementation. Instructional Science, 1976, 5, 343-364.

Leinhardt, G. Evaluating an adaptive education program: Implementation to replication. Instructional Science, 1977, 6, 223-257. (a)

Leinhardt, G. Program evaluation: An empirical study of individualized instruction. American Education Research Journal, 1977, 14, 177-293. (b)

Lohnes, P. R., & Cooley, W. W. Regarding criticisms of commonality analysis. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, March 1978.

MacDonald, F. J., & Elias, P. Beginning teacher evaluation study (Phase II, Technical Summary, Final Report). Princeton, N.J.: Educational Testing Service, 1976.

APPENDIX

Rationale For The Constructs, Variables, And Measures Included in IDS

Rationale for Opportunity Construct

This construct refers to the opportunity that the students have to learn criterion-relevant material. There are two variables in the Opportunity construct: time and curriculum. Time refers to the amount of access that a student could have to criterion material. Curriculum refers to the relationship between the instructional material presented and the criterion, which is the CTBS. The amount of time available to learn criterion-relevant materials consists of eight measures. The curriculum opportunity measure consists of two measures. This rationale describes each of the measures, defines them, describes if there is a theory or general hypothesis about whether the measures would be expected to affect achievement positively or negatively, describes why they are in the variables, and describes any ambiguities that exist.

Time

1. Number of students that are enrolled refers to the number of students enrolled in the classroom on any given day. Theoretically, if the number of adults in the room remains constant, the more students enrolled the less opportunity any one child would have to be exposed to material. Teachers generally believe that the larger a class, the harder it is for them to be successful as teachers. Therefore, we expect the number of

students enrolled to be related negatively to overall classroom achievement.

2. Percent of students present refers to the number of students present on a given day, divided by the number enrolled. Given that a student is enrolled, the more that the student is present, the greater the opportunity the student has to learn the material. This measure is included in the time variable because it is a reflection, at the classroom level, of the number of days that the student has available to learn material.

3. Number of minutes per week in the subject matter refers to total minutes per week spent in either the reading or mathematics curricula as reported by the teacher. Obviously, the more time that is spent in a given subject matter area, the greater the opportunity for the students to learn the appropriate material.

4. Number of students who transfer in refers to the number of students who come into the classroom after the start of the school year. In general, it is felt that the more movement in and out (see next measure) of a room, the less opportunity students have to learn. These disruptions negatively affect both the transient students and the permanent ones.

5. Number of students who transfer out refers to the number of students who leave the classroom before the end of the school year. (See the rationale given in the measure above.)

6. Regular Instruction Time is an estimate of the time by year allocated to the subject matter. The more amount of time allocated the greater the opportunity to learn.

7. Percent of students on-task refers to how many students in a class are actually engaged in instructional activities at a given time. This, in combination with the percentage of students present and the amount of time allocated per week, is the closest estimate we have of how much instruction a student actually receives. Instruction received is clearly a function of how much time is available and how much that time is utilized. It may be that two classrooms have identical quantities of time available, but very different utilizations of that time in terms of how many students are actually on-task. We would expect a positive relationship between the percentage of students on-task and achievement.

8. Days present is an estimate of the proportion of days present between the pretest and the posttest. The more time present in school the greater the opportunity to learn the criterion material.

9. Days absent is an estimate of the average number of days absent per student. The more days absent there are the less opportunity to learn there is.

Curriculum

1. Overlap estimate between curriculum and criterion test as assessed by the teacher refers to the degree to which the items on a criterion test represented the kinds of skills that were included in the instruction for a given classroom, as estimated by the teacher. This measure is expected to have a positive correlation with student achievement because the more the test reflects what has been taught, the greater the chance that the individual student can perform well on that criterion test.

2. Overlap estimate between curriculum and criterion test as assessed by the curriculum analyst reflects an estimate of overlap of relevant criterion test items based only on the material present in the curriculum. This estimate is included in addition to the teacher's estimate because the teacher's estimate could be biased by what he or she presumes the student knows; the curriculum analyst's estimate reflects only materials actually in the curriculum text. The combination of both measures should be a reasonably powerful predictor of achievement as regards learning opportunity.

Rationale for the Motivators Construct

The Motivator's construct consists of two variables, Curricular Motivators and Interpersonal Motivators. The first variable is related to aspects of the curriculum that support and encourage student learning, and the second variable is related to those aspects of teacher behavior, other than explicit instruction, that support learning. While there has been considerable research on Motivation as a construct, there has been limited research on validating exactly which elements in a classroom motivate students to learn. There are 10 measures nested within the two variables. All of the measures in the Motivator construct are entered positively.

Curricular Motivators

1. Degree of interest of curricular materials in general refers to how well the curricular materials capture and hold a student's interest. The more interesting the learning material is, the more one would expect a student to desire or to exhibit the desire to interact with that material. Obviously, deciding what is interesting and what is not interesting is a highly subjective act, but once having established some guidelines, the more interesting material is more likely to motivate the student to learn.

2. Number of modes of instruction refers to the modality, such as videotapes, audiotapes, and language masters, in which instruction is presented as estimated by a curriculum analyst. We assume that it is motivating to have different modes departing from the traditional textbook situation, we are not positive that it does more than spark interest. The hope is that the more modes there are, the more chances one has to meet specific student needs for a particular modality.

3. Speed of correction refers to the speed with which students have their work products corrected as estimated by the teacher. Learning theory informs us that the more quickly a student receives corrected materials, the more likely that the feedback will be informative. However, the speeds to which the psychologists refer are usually of a minute or less. The speeds in which instructional environments operate are usually of a week or less. We assume that it is still more motivating to have work returned promptly than to wait for it. This measure was assigned to the curricular motivators because it was felt that speed of correction was largely a product of the curriculum rather than an interpersonal experience. That is, the student performed on curricular materials and received the feedback in a written mode or occasionally in a verbal mode.

Interpersonal Motivators

1. Presence of peer tutoring has been shown to be an effective instructional device, but it is not clear whether it is effective because it is highly motivating, or because the actual

instruction received is more precise or more useful. Apparently, peer tutoring works because it is fun, rather than because the instructional content differs from that given by the teacher.

2. Use of games and contests in the classroom refers to games and contests used in relationship to instructional material, or as a direct reinforcer for having mastered certain units of material. Obviously a highly interpersonal activity, it is included in this set of motivators.

3. Self-evaluation refers to the right of the student to evaluate him or herself, either by using a scoring key, or some other means by which students can assess their own work without teacher intervention. It is included as an interpersonal motivator because it is an aspect of controlling the interpersonal situation, as opposed to a known instructional technique. The ability to be one's own motivator seems to affect learning in a positive fashion.

4. Students choose their own seating relates to one aspect of student self-management, the ability to control one's own interpersonal environment. As such, it (and the measure that follows) is an indirect motivator, but is related to an interpersonal type of motivation because it reflects how the child decides whether or not he or she will be near peers or away from them.

5. Students choose to work alone is another aspect of student self-management. The rationale is given above.

6. Degree of teacher contingent praise refers to praise given to a student or students contingent on work-related behaviors. This is considered a more important motivator than general praise (see next measure). It is expected that teacher praise, being positive and supportive of learning behaviors, should increase the probability that the students will engage in those behaviors.

7. Degree of teacher general praise refers to praise given for positive behaviors other than work-related ones. The rationale for inclusion is given above. These measures are clearly interpersonal in that they involve the teacher directing verbal behaviors towards the students.

Rationale for Instructional Events Construct

Instructional Events refer to the actual teaching behaviors that occur in the classroom. The five variables that make up the Instructional Events construct are: Management Statements, Cognitive Instruction to Individuals, Cognitive Instruction to Small Groups, Cognitive Instruction to Large Groups, and the Quality of Instruction in general. There are 29 measures that compose these five variables. The following rationales describe the measures within each variable. All but two of the measures in the Instructional Events construct are combined positively.

Management Statements

1. Management statements to individuals refers to statements made by a teacher that are directed to an individual child, and that focus on managerial rather than cognitive information (such as telling the child to go to a certain place in the classroom, behave in a certain way, or procedural aspects of behavior in the class). In order to teach in any program, a certain amount of management-type information must be communicated to the student. However, if management information is greater than the absolute amount needed, or if a specific program requires a tremendous amount of management information, this information and the time spent communicating it will be taken away from cognitive information and so tend to depress the actual instruction received.

2. Management statements to groups refers to the same type of managerial statement made by the teacher, but it is directed to the entire class or to groups of students within the class. It is typified by general instructions for changing periods or leaving the classroom, or a request for quieting down when the entire classroom is involved. These two management statement measures would be expected to relate negatively to achievement but positively to each other.

3. Cognitive management statements to individuals refers to teacher statements that are not instructional in nature, but are directly supportive of cognitive behaviors, such as telling the child how many assignments will be given for the week. However,

the same problem exists with cognitive management statements. Cognitive management statements are necessary, but when they are excessive, they detract from instruction.

4. Cognitive management statements to groups refers to teacher statements of this nature that are made when the entire class is assembled, and the students are told to do one activity or another. This entire variable is expected to be related negatively to achievement; that is, a classroom would be considered successful if it had minimal management information.

Cognitive Instruction to Individuals

1. Cognitive statements to individuals refers to those cognitive statements, directed at individual students, that do not occur with any other type of interaction. A statement is distinguished from a question in that it does expect and does not receive a response from a child. This is common when the teacher is giving an explanation, or tutoring. It is generally better to give cognitive statements than it is to give managerial statements, but it is better to ask questions than it is to make statements. Statements do not provide the child with any chance to practice. However, it is added as a positive measure.

2. Cognitive questions to individuals refers to teacher statements that elicit a student response of some type. The child may make a response verbally or in writing. In general, verbal responses are expected. The more of these that occur, the

more a child has the opportunity to learn the subject matter of interest.

3. Individual child response refers to student responses to questions, although occasionally they can be delayed responses to statements. This category refers only to those cases in which one child is answering the question.

4. Ratio of cognitive statements to managerial statements asssumes that the higher the ratio of cognitive statements to management statements the better. The reason for having this measure is that it helps control for some of the teachers who talk a lot, both managerially and cognitively. It is possible for a teacher who does a lot of instruction also to make many management statements, and for a teacher who talks very little to make many (or few) management statements.

5. The number of cognitive contacts refers to such contacts made to different children. This permits an assessment of the rate of contact, which is assumed to facilitate learning.

6. Tutoring individuals refers to the rate of tutoring given in excess of a predefined time limit. Tutoring is complicated. It is obviously a very good thing to do when an individual child actually needs it, but teachers often engage in extended contacts because they are more reinforcing, essentially ignoring other children. Thus, while tutorial contacts are good in one sense, they should be limited. However, this measure is included positively.

Cognitive Instruction to Small Groups

1. Cognitive statements to small groups of children refers to cognitive statements directed to a small group rather than to an individual child within that group.

2. Cognitive questions to small groups of children refers to cognitive questions directed to a small group as a whole.

3. Children responses to cognitive questions refers to responses that are choral, or at least from more than one child in the small group. If, for example, a teacher is working with a group of children in a reading lesson, asks a question, and receives a group response, it would be counted here.

4. Number of cognitive contacts refers to such contacts made by the teacher to the small group of children as a whole.

5. Frequency of personal or tangential information refers to such contacts made to either individuals or to groups. It has been placed in the small group category because the most frequent occurrence of personal or tangential information is within the small group setting. These two variables, instruction to individuals and instruction to small groups, are closely related to the Rosenshine variables of direct instruction.

Cognitive Instruction to Large Groups

1. Cognitive statements refer to those directed to the entire class, in general, or to very large subsets of the class.

The measures are otherwise parallel in form to contacts and responses for individuals and small groups.

2. Cognitive questions. See above.

3. Amount of lecture time refers to time in which the teacher alone is talking, when the time is in excess of a specified amount. This is expected to have a negative influence on learning, and the measure is subtracted from the variable. Thus, the large groups instruction variable consists of making cognitive contacts with students and permitting the students to respond, and minimizing the number of direct cognitive statements made by the teacher.

Quality of Instruction

1. Silence refers to teacher nonverbal behavior and can be thought of as "down" time. For example, it is the case in which the teacher is standing in the corner of the room talking to another adult, or the teacher has stopped instructing and nobody else is doing anything in the room. The measure is weighted negatively in creating the quality of instruction variable.

2. Degree to which active responses are sought refers to the teacher's attempts to have a child or children participate and respond actively in learning events. This may be done in ways that are not picked up by the frequencies measured in the earlier variables.

3. Degree to which the teacher refers to earlier curricular material is considered an effective teaching strategy, as it links the child with materials that presumably have already been mastered, encouraging the child to see the relationships between what has been taught and what is being learned now.

4. Degree to which the teacher focuses the child's attention refers to the teacher's efforts (when teaching) to have the child's attention on the task rather than on another child, etc. Thus, the more the teacher is directing the child's attention to the task, the better the teaching.

5. Degree to which the teacher solicits the child's opinion on the correctness of the response is included because it indicates that the teacher is making the child practice a type of self-evaluation skill.

6. Degree to which the teacher refers to earlier success refers to earlier successes the child has had in dealing with a similar type of problem. This is a type of reinforcement and encouragement to face a new and perhaps more difficult instructional task.

7. Degree to which the teacher tutors effectively refers to how clearly and precisely the extra instruction is given, and how well the teacher gives examples and elicits counter-examples from the child.

8. Degree to which the teacher manages the class effectively refers to the success the teacher has in managing the class. This is the opposite of a large number of management statements. If a teacher has a class well under control, even if the class style is very open and flexible, the teacher does not have to give a large number of managerial statements, the teacher does not have to correct bad behaviors, and the children are not constantly seeking management-type information.

9. Degree to which the child receives feedback refers to the degree to which the teacher gives the child pertinent feedback on performance as soon as that performance occurs.

10. Degree to which the teacher contacts individual students refers to whether the teacher is generally contacting individuals or is generally contacting groups. Presumably, frequent teacher contacts facilitate learning.

Rationale for the Structure Construct

The Structure construct refers to the organization, presentation, and utilization of the curricula. The six variables that make up the Structure construct are: Placement and Pretest Procedures, Mastery Assessment Procedures, Matching Practices, Grouping and Assignments, Instruction for Special Needs, and Sequencing and Pacing. There are 45 measures nested within these six variables, five of which are added negatively.

Placement and Pretest Procedures

1. Curricular placement procedures refers to the curriculum analyst's assessment of whether there are any placement procedures available in the curriculum as provided by the publisher. It is combined positively.

2. Standardized tests for initial placement is obtained by asking the teachers how they place the students. One would assume that the use of standardized tests is less desirable than a criterion-based test for placement, but more desirable than no placement procedure at all. Therefore, the use of it is included in a positive direction with the other initial placement procedures.

3. Teacher uses curriculum provided tests for initial placement. Unlike the first measure, this is not simply whether the curriculum has them, but whether the teacher uses them to place students. It, too, is combined positively.

4. Teacher uses teacher-made tests for initial placement assumes that teacher-made tests would closely reflect what is in the curriculum. Presumably, if a teacher uses teacher-made tests to place students, it would be better than not using anything, and might be better than using curriculum-embedded tests if the teacher deviates from curricular material in instructing. It is combined positively.

5. Teacher uses observation of a work sample for initial placement is similar to the use of a teacher-made tests and curriculum-embedded tests in many individualized curricula. Rather than structuring a testing situation, the teacher uses work products to indicate the child's level of competence and to determine whether the child should begin instruction.

6. Curricular pretests refers to the curriculum analyst's assessment of the presence or absence of pretests available in the curriculum, as provided by the publisher. The distinction between pretests and placement procedures has become a little vague, but insofar as there are slight differences between placement procedures and pretests, both are included in this variable.

Mastery Assessment Procedures

1. Curricular posttests refer to the curriculum analyst's assessment of whether there were any posttests available in the curriculum as provided by the publishers.

2. Presence and use of some curriculum-provided method to assess mastery adds to the previous measure information about whether the teacher actually uses a curricular method to assess mastery.

The following measures (3 through 8) are alternative ways in which a teacher can check on student progress. They are all added psoitively. However, they become less and less specific and rigorous, because there is decreasing distinction between

instruction and assessment. This is a common problem in individualized instruction, where teachers are quite sure a student knows something and are not willing to accept the inability of a student to perform in an objective or non-teaching situation. One would assume that the best way of assessing instruction would be associated with curriculum-embedded tests or teacher's tests, or a specific class assignment that was given. As the range of assessment procedures moves toward talking to the students about their work, one has a less accurate assessment procedure.

3. Teacher uses a curriculum-provided test to assess mastery is one way for a teacher to assess mastery, particularly where the teacher has adhered closely to the curricular sequences and materials.

4. Teacher uses a teacher-made test to assess mastery is an alternative for assessing mastery that is particularly appropriate where the teacher has substantially altered the curriculum.

5. Teacher reviews in-class assignments to assess mastery is parallel to the use of work samples to place students. It is an informal but potentially informative method of assessing student knowledge.

6. Teacher uses the completion of a specified work unit to assess mastery, while it is not really a way of assessing

mastery, assumes that if a student reaches a place in the curriculum, then he or she knows the material in it.

7. Teacher observes a work sample to assess mastery may be similar to spot checking, although it is not a formal mastery procedure.

8. Teacher confers with the pupil to assess mastery parallel to an oral quiz and may be quite effective.

Matching Practices

1. Number of days since last test was given assumes that the longer the time between tests, the less knowledge a teacher has about what the student actually knows, and the less likely he or she is to adjust instruction to meet the student's needs. This measure goes into the matching practices variable negatively.

2. Range of days since last test was given assumes that the greater the range of days across students, the less likely it is that students are being regularly instructed and regularly assessed on their progress. This measure goes into the variable negatively.

3. Percent correct on last test assumes that if students are tested regularly on material that they have been well taught, a high percentage should do well on tests; if on the other hand, tests are used to punish or control or are not tied to instruction, only the "bright" students will do well. High

percentages of correctness indicate a good fit between test and instruction.

4. Frequency of testing is an estimated rating to assess information similar to 1 and 2. It is added positively.

5. Use of tests for purposes of assessing mastery is an indication that the teacher is using tests to gain information about student knowledge levels, after instruction.

6. Use of tests for purposes of diagnosis is an indication that the teacher is using tests to gain information about student knowledge levels, either prior to instruction, or when difficulties occur.

7. Use of tests for purposes of prescription indicates that teachers use errors on tests to guide them in making assignments for students.

Grouping and Assignments

1. Percentage of unique assignments refers to the number of different assignments made to individuals on a given day. The more unique the assignments are, the more individualized the actual instruction received will be. The measure is combined positively.

2. To whom the assignments are made, individuals or groups, assumes that the more asssignments that are made at the student level (rather than at the group level or at the entire class

level), the more individualized the instruction is. The measure is combined positively.

3. Percentage of time spent in group of entire class refers to instruction given in the class as a whole. This measure is an estimate of non-individualization and so is put into the variable negatively.

4. Percentage of time students spend in subgroups of 11 or more is large group instruction and is subtracted from the variable.

It should be remembered, however, that the entire variable, Assignment and Grouping, may be positively or negatively related to achievement. Although we arbitrarily decided to call individualization the positive end of this scale, and large group instruction the negative end, if large group instruction is important, this will be reflected by a negative correlation between this variable and the criterion.

5. Percentage of time students spend in subgroups of two to four, or in small group instruction, is included in the variable positively, as individualized instruction.

6. Percentage of time students spend in one-to-one instruction is individualized teaching. It is added positively.

7. Number of students in the group is negatively related to individualization, since, presumably, the fewer students in the group, the more individualized the instruction.

8. Frequency of regrouping assumes that even if large groups of students are working together, frequent regrouping may be a way of rematching them with the curriculum. It is added positively. Because assignment procedures closely follow grouping procedures, the unit of assignment is almost always the same as the unit of grouping.

Instruction for Special Needs

1. The curriculum suggests additional practice exercises when difficulty occurs, is clearly one alternative way in which the curriculum and the teacher are responsive to specific, unique needs that occur in the course of instruction.

2. The curriculum suggests repeating incompletely mastered materials when difficulty occurs, is not as individualizing or as unique a way of meeting this need. It is preferable for new material to be provided that covers the same area, rather than sending a student back through exactly the same material. However, the measure is added positively.

3. Teacher has child repeat content using new materials is presumably an appropriate strategy for helping a student who is having difficulty with instruction.

4. Teacher gives one-to-one remedial instruction when difficulty occurs, is especially appropriate when the difficulty arises because the student does not understand a specific element. It is combined positively. It is not appropriate if the student consistently needs help.

5. Teacher uses curriculum-suggested remedial material is appropriate when the difficulty is one that is covered in the remedial material provided.

6. Teacher continues on with new material, despite difficulties, with the intention of returning to the incompletely mastered material later. This approach would probably be reasonable in reading, but not as reasonable in math, where there is more of a natural hierarchy. However, the measure is combined positively.

7. Readability level of curriculum material is only estimated for the reading material. We felt it belonged in this set because it reflected the degree to which group needs were g met; that is, if the material is inappropriate for the entire lass, it cannot possibly meet individual needs. The more appro.ate the readin evel, the more ndividual the needs are being t.

Sequencing and pacing

1. Clarity of the sequence of instruction refers to how clear the progression of materials is to the teacher and student.

Obviously, the clearer the sequence is, the easier it should be to use the materials and gain from them.

2. Extent to which the teacher follows the sequence. Presumably, if the sequence(s) is (are) clear, the teacher should follow it. If the teacher departs from following it, it may result in a less clear presentation to the student.

3. Extent to which the sequence varies within the classroom is a positive type of activity depending on the content of the material and the general flexibility of the program. Presumably, the more a teacher is able to vary a sequence within a classroom, the more individualized it is.

4. Extent to which the teacher creates supplemental materials to improve sequencing. The value of this activity depends on how necessary it actually is. Creating new materials and slowing the class down may not be a particularly good thing to do if, in fact, the students do not need that changed in the sequence. However, when needed, it is positive, and the measure is combined positively.

5. Each segment of instruction is prerequisite to the next. This depends on whether the segments of instruction are absolutely prerequisite or not to each other. It is a good situation when the subject matter is also tightly related, it is not as useful when this is not the case. The measure is combined positively.

6. Different sequences are given based on a student's preference. Obviously, this is something that should only be used to improve sequencing when the sequencing is relatively independent of the subject matter hierarchy. The measures in sequence are also related to pacing. The clearer the sequence, the easier it is for varying rates of pacing to take place.

7. Presence of self-pacing. The more self-pacing there is, the more individualized the instruction.

8. Presence or absence of rules and guidelines for pacing. It would seem necessary for some type of guidelines to occur in order for self-pacing to work.

9. Amount of reviewing of instruction. The more reviewing that occurs the slower the pace will be. However, more reviewing may increase the level of mastery and retention. Therefore, amount of reviewing is included as a positive measure, even though it is not expected to be closely related to a fast pace in going through the curriculum.

Table 1

Number of IDS Classrooms Included in these Analyses

GRADE 1

MATH \ READING	Included	Out	
Included	51	35	86
Out	53	61	114
	104	96	200

GRADE 3

MATH \ READING	Included	Out	
Included	57	62	119
Out	52	29	81
	109	91	200

Table 2

CTBS Correlations

Grade 1 Reading		Pre	Post	Gain
N=104	Pretest	1.00	.50	-.13
	Posttest		1.00	.79
	Gain			1.00

Grade 1 Math		Pre	Post	Gain
N=86	Pretest	1.00	.39	-.18
	Posttest		1.00	.84
	Gain			1.00

Grade 3 Reading		Pre	Post	Gain
N=109	Pretest	1.00	.86	-.12
	Posttest		1.00	.41
	Gain			1.00

Grade 3 Math		Pre	Post	Gain
N=119	Pretest	1.00	.78	.04
	Posttest		1.00	.66
	Gain			1.00

Table 3

OPPORTUNITY

Variable	#	Measure	n=104 GR1 $\overline{X}$	Read S.D.	n=86 GR1 $\overline{X}$	Math S.D.	n=109 GR3 $\overline{X}$	Read S.D.	n=119 GR3 $\overline{X}$	Math S.D.
Time	1	# student enrolled	25.52	2.60	25.81	3.11	26.41	3.78	26.28	3.78
	2	% student present	92.76	5.63	93.32	6.15	94.39	6.27	94.48	5.66
	5	# min./wk. in subj.	558.51	135.80	247.36	70.40	454.54	117.20	266.41	78.93
	7	# stud. transfer in	3.39	2.17	3.81	2.44	2.84	2.30	2.99	2.69
	8	# stud. transfer out	4.48	3.23	4.81	3.12	3.73	2.91	3.79	2.87
	12	% student on task	59.64	12.43	56.35	14.97	57.15	14.96	55.29	16.28
	238	Reg. Instruc. Time	15022.74	3678.16	6629.59	1911.16	12213.04	3198.34	7087.08	2247.54
	239	Days Present	134.50	4.34	133.84	4.64	134.26	3.94	134.71	4.02
	240	Days Absent	10.76	3.37	11.02	3.98	8.76	2.44	8.41	2.19
Curriculum	13	overlap-teacher est.	50.93	18.46	27.59	12.33	51.12	21.33	56.14	19.40
	14	overlap-curr. anal.	27.13	14.91	15.03	10.38	20.02	5.60	30.52	14.56

Deleted measures:	Reason:
3 - # minutes in school day	Subject better measure
4 - # person hrs./wk. parent/aide instructors	Poorly measured, 60% missing data or zeros
6 - # min./night homework	Poorly measured
9 - average # days attendance	Small variance, poorly measured
10 - # days in school year	No variance, 74% report 180 days
11 - # days Bet. pre and post	Missing Data

Table 4

MOTIVATORS

Variable	# Measure	n=104 GR1 $\overline{X}$	Read S. D.	n=86 GR1 $\overline{X}$	Math S. D.	n=109 GR3 $\overline{X}$	Read S. D.	n=119 GR3 $\overline{X}$	Math S. D.
Curriculum	15 interest in curr. mat.	4.51	.58	3.99	.82	4.27	.61	3.85	.74
	22 # modes of inst. - A	9.37	4.23	5.60	3.34	8.39	4.28	7.56	3.45
	24 speed of correction	3.53	.73	3.66	.80	3.01	.87	2.98	.74
Interpersonal	25 peer tutoring	3.33	.81	3.17	.83	3.22	.68	3.21	.68
	26 use of games/contests	3.15	.96	3.00	.96	2.53	1.07	2.78	.90
	27 self-evaluation	.47	.87	.56	.89	.84	1.17	1.17	1.19
	28 self-manag.-seating	3.72	1.43	2.26	1.78	3.12	1.62	2.20	1.66
	29 self-manag.-work alone	2.06	1.43	1.79	1.47	2.15	1.35	2.12	1.29
	30 teacher contingent praise	19.33	14.42	17.67	14.03	12.98	10.32	15.63	12.40
	31 teacher general praise	28.80	13.23	27.50	13.21	22.94	11.57	22.90	11.15

Deleted measures:	Reasons:
16 - Interest in teacher's materials	Variance close to zero
17 - Interest in student workbook	Non-useable data with large number of out of range values
18 - Interest in student workbook	Non-useable data with large number of out of range values
19 - Variation in teacher's materials	Variance close to zero
20 - Variation in student textbook	Non-useable data with large number of out of range values
21 - Variation in student workbook	Non-useable data with large number of out of range values
23 - # modes of instruction-B (interview)	Confusing with measure 22
32 - Teacher negatives to individuals	Variance close to zero
33 - Teacher negatives to groups	Variance close to zero

Table 5

INSTRUCTIONAL EVENTS

Variable	#	Measure	n=104 GR1 $\overline{X}$	n=104 Read S. D.	n=86 GR1 $\overline{X}$	n=86 Math S. D.	n=109 GR3 $\overline{X}$	n=109 Read S. D.	n=119 GR3 $\overline{X}$	n=119 Math S. D.
Management	34	indiv. manag. statements	22.90	13.63	27.66	14.78	18.53	11.13	21.89	13.24
	35	group manag. statements	11.18	6.43	15.89	9.92	8.62	4.87	10.07	9.07
	36	indiv. cog. man. stmts.	23.34	13.37	32.66	14.92	27.22	16.20	37.04	20.21
	37	group cog. man. stmts.	20.41	10.04	21.76	8.85	17.11	8.31	20.45	11.83
Individual	38	indiv. cognitive statements	8.41	5.74	12.52	8.82	10.22	7.97	15.68	11.23
	40	indiv. cognitive questions	54.16	23.08	52.65	20.98	51.32	22.46	54.02	21.81
	43	indiv. child responses	78.17	24.16	68.73	22.41	79.99	25.79	71.70	27.75
	45	ratio cog. to management	2.18	1.19	1.55	.82	2.25	1.30	1.89	1.46
	46	indiv. cognitive contacts	27.85	12.76	27.19	13.83	26.81	11.97	25.67	10.44
	54	extended tutor time	5.13	4.53	4.75	4.24	7.56	8.27	7.75	6.39
Small Group	39	sm. gr. cognitive stmts.	14.65	8.26	7.47	7.45	15.39	10.61	9.97	13.54
	41	sm. gr. cognitive questions	56.86	23.76	25.80	24.74	47.20	25.46	22.82	21.70
	44	sm. gr. child responses	45.99	23.45	22.61	22.14	35.43	24.73	17.75	19.90
	47	sm. gr. cotnitive contacts	22.26	8.09	18.30	13.63	20.04	11.31	17.02	18.36
	52	personal/tangential info.	3.67	3.82	1.61	2.97	3.42	4.27	1.64	2.52
Whole Class	48	whole - cognitive stmts	2.43	4.79	9.64	10.68	2.50	4.66	8.63	10.73
	49	whole - cognitive questions	6.48	10.80	23.19	20.85	6.22	9.83	20.37	22.38
	50	whole - child responses	5.22	9.88	22.92	22.04	4.38	7.68	18.42	20.02
	55	lecture time	.82	1.14	.76	1.15	.88	1.29	1.00	1.54
Quality	51	silence/non-verbal time	7.09	5.36	7.94	6.74	8.96	7.12	8.42	7.51
	59	active response sought	43.41	6.28	41.51	8.05	39.95	7.74	38.28	9.01
	61	refer to earlier curr. info.	9.47	8.03	7.97	7.33	11.33	7.75	10.88	8.28
	62	focuses child attention	39.81	8.42	37.44	9.20	38.35	7.94	36.18	9.20
	63	solicit opinion on correct.	12.50	10.52	16.11	13.29	13.39	10.69	16.77	12.26
	64	refer to earlier success	2.64	4.72	2.79	4.83	2.62	5.12	3.03	6.77
	65	teacher tutors effectively	42.98	7.33	41.40	8.21	41.42	7.67	41.35	7.41
	66	teacher manages effectively	40.72	8.60	39.30	9.18	39.63	8.07	39.50	9.28
	67	child receives feedback	28.89	12.60	28.14	10.63	25.78	10.46	28.95	10.87
	68	teacher interacts w/ indiv.	36.01	9.61	36.45	9.30	34.82	9.67	34.92	8.71

Deleted measures:	Reason
42 child - initiated responses - individual	43% zeros
53 teacher errors	almost no occurrence
56 # years teaching experience	poorly measured
57 teacher rating of amt. of training received	poorly measured
58 teacher rating of quality of training received	poorly measured
60 teacher models responses	poorly measured

Table 6

STRUCTURE

Variable	#	Measure	n=109 GR1 $\bar{X}$	n=109 Read S. D.	n=86 GR1 $\bar{X}$	n=86 Math S. D.	n=109 GR3 $\bar{X}$	n=109 Read S. D.	n=119 GR3 $\bar{X}$	n=119 Math S. D.
Placement & Pretest	69	curr. ref. placement test	1.32	.35	1.47	.45	1.32	.37	1.49	.46
	70	place. proc.-standard. test	1.36	.48	1.22	.42	1.25	.43	1.11	.31
	71	place. proc.-curr. test	1.33	.47	1.23	.43	1.47	.50	1.37	.49
	72	place. proc.-teacher test	1.43	.50	1.52	.50	1.27	.44	1.50	.50
	74	place. proc.-obs. wk. sample	1.72	.45	1.65	.48	1.58	.50	1.43	.50
	79	curric. pretests	1.03	.09	1.28	.44	1.01	.03	1.59	.47
Mastery	80	curric. posttests	1.25	.32	1.68	.40	1.41	.36	1.91	.24
	81	curric. meth. assess. mastery	1.98	.11	1.86	.28	1.96	.15	1.84	.31
	84	teach. assess.-curr. test	1.77	.32	1.58	.38	1.85	.28	1.63	.40
	85	teach. assess.-teacher test	1.61	.40	1.74	.37	1.57	.41	1.80	.33
	86	teach. assess. - in-class assign.	1.41	.38	1.47	.39	1.47	.37	1.46	.38
	88	teach. assess.-compl. wk. unit	1.34	.34	1.29	.33	1.37	.38	1.36	.39
	89	teach. assess.-observe work	1.73	.36	1.70	.35	1.73	.35	1.59	.38
	91	teach. assess.-conf. w/ pupil	1.17	.27	1.15	.28	1.25	.33	1.22	.33
Matching Practices	110	# days since last test	14.88	12.47	10.17	7.98	13.21	12.34	10.12	13.15
	111	range days since last test	16.93	20.80	5.67	12.22	17.64	22.91	9.76	21.46
	112	% correct last test	85.15	12.03	84.72	11.21	82.73	10.51	81.77	11.49
	113	frequency of testing	2.92	.81	2.88	.80	3.04	.87	3.27	.85
	116	tests for assess mastery	1.85	.25	1.84	.26	1.89	.26	1.92	.23
	117	tests for diagnosis	1.84	.27	1.85	.29	1.80	.31	1.76	.33
	118	tests for prescription	1.64	.37	1.63	.39	1.60	.40	1.59	.41
Grouping & Assignment	108	% unique assignments	21.01	18.65	10.15	12.22	25.95	25.82	20.18	27.61
	109	to whom assign. made	2.14	.57	1.90	.61	2.20	.56	2.20	.88
	134	% time - entire class	15.67	12.39	34.15	26.25	12.05	11.40	27.68	24.69
	135	% time - subgr. 11+	8.98	13.69	17.64	21.53	15.11	18.57	17.83	21.18
	137	% time - subgr. 2-4	10.42	9.34	6.88	6.83	11.06	10.23	9.45	10.95
	138	% time - 1 to 1	13.02	9.41	17.05	14.22	15.72	17.72	23.28	22.02
	141	# students in group	7.61	2.85	15.12	6.13	9.23	5.79	15.83	6.99
	142	frequency of regrouping	3.33	1.83	1.89	2.05	2.17	1.87	2.22	2.46

Table 6 (continued)

STRUCTURE

Variable		GR1 Read $\bar{X}$	GR1 Read S. D.	GR1 Math $\bar{X}$	GR1 Math S. D.	GR3 Read $\bar{X}$	GR3 Read S. D.	GR3 Math $\bar{X}$	GR3 Math S. D.
Meeting	94 diff.-curr.-addn. practice	1.46	.35	1.65	.44	1.41	.35	1.65	.43
Individual	95 diff.-curr.-repeat incomplete	1.59	.38	1.33	.44	1.53	.41	1.65	.42
Needs	97 teacher alt.-repeat material	1.64	.37	1.62	.41	1.63	.36	1.55	.40
	101 teacher alt.-1 to 1 remedial	1.66	.38	1.66	.34	1.73	.36	1.80	.30
	102 teacher alt.-curr. sugg. remed.	1.29	.36	1.25	.36	1.26	.34	1.22	.32
	104 teacher alt.-cont. on, back later	1.37	.41	1.36	.43	1.37	.38	1.34	.39
	145 readability level	14.32	5.32	--	--	23.56	3.56	--	--
Sequence	121 clarity of sequence	3.81	.79	3.32	.85	3.70	.72	3.52	.86
	123 teacher follows sequence	1.89	.78	2.24	.90	1.88	.77	2.31	.92
	124 sequence varies w/in classroom	.89	.92	.95	.96	.94	.93	.93	1.00
	125 teach. create suppl. to imp. seq.	3.04	.98	3.24	.94	3.07	.90	3.21	.88
	126 ea. seg. instr. prereq. to next	1.12	.24	1.14	.31	1.08	.17	1.11	.29
	127 diff. alt. based on stud. pref.	1.02	.08	1.00	.04	1.01	.03	1.15	.32
	130 self-pacing	4.79	5.88	3.55	5.32	4.45	5.53	5.18	5.90
	132 rules/guide. for pacing	1.10	.25	2.12	.84	1.06	.27	1.92	.90
	133 amount of review	1.05	2.35	7.24	15.46	.19	.82	9.94	6.65

deleted measures:	Reason:
73 placement procedure - placement at end of previous year	illogical for grade 1, not a placement method
75 placement procedure - conference with pupil	no variance
76 placement procedure - start at beginning	not placement procedure, little variance
77 placement procedure - conference with specialist	no variance
78 placement procedure - wishes of child/parent	infrequent occurrence
82 presence of curric. method to assess overall learning	too similar to 81
83 teacher assess. mastery - standard. test	infrequent occurrence
87 teacher assess. mastery - review homework	infrequent occurrence
90 teacher assess. mastery - conference with specialist	infrequent occurrence
92 teacher assess. mastery - other	all "other" dropped
93 difficulty encountered - curriculum - no additional practice	not a method of helping
96 difficulty encountered - curriculum - refer to other material	all "other" dropped
98 teacher alternative - repeat identical	infrequent occurrence
99 teacher alternative - give extra homework	infrequent occurrence
100 teacher alternative - extra in-class assignment	infrequent occurrence
103 teacher alternative - depends on child	infrequent occurrence, non specific alternative
105 teacher alternative - consult with specialist	infrequent occurrence

Table 6 (continued)

STRUCTURE

deleted measures:	Reason:
106 teacher alternative nothing	0 variance
107 teacher alternative - other	all "other" deleted
114 degree testing varies by student	is measured better in 108
115 tests for grading	
119 tests used - other	all "other" deleted
120 overlap between mastery tests & curriculum	no variance
122 presence of explicit rules for sequencing	no variance
128 materials may be used in any order	no variance
129 no explicit information on sequencing given	opposit of 122
131 extent to which students must meet deadlines	70% zeros
136 % time spent in groups of 5-10	information contained in other retained measures
139 % time spent in other	its an "other"
140 # of groups	information in other measure
143 specificity of objectives	variable deleted
144 degree to which materials match objectives	no variance

Table 7

Outcome Variance Explained
(18 Predictors)

	Number of Classrooms	R	R^2	Adjusted R^2
Grade 1 Reading	104	.71	.51	.41
Grade 1 Mathematics	86	.66	.43	.29
Grade 3 Reading	109	.90	.81	.77
Grade 3 Mathematics	119	.85	.73	.68

Table 8

Commonality Results for Grade 1 Reading
(N=104 Classrooms)

	Construct	Uniqueness	Commonalities 12	13	14	24	25	Prob.
1.	Pretest	.06	-.03	.02	-.01			.02
2.	Opportunity	.22	-.03			.01	.03	.0001
3.	Motivators	.01		.02				.49
4.	Instructional Events	.01			-.01	.01		.92
5.	Structure	.02					.03	.87
6.	Setting	.00						.87

Total $R^2 = .36$, $P < .001$
Total Process Uniqueness = .27

Table 9

Commonality Results for Grade 1 Mathematics
(N=86 Classrooms)

	Construct	Uniqueness	Commonalities 13	Commonalities 25	Prob.
1.	Pretest	.09	.04		.01
2.	Opportunity	.09		.02	.01
3.	Motivators	.03	.04		.25
4.	Instructional Events	.04			.61
5.	Structure	.06		.02	.40
6.	Setting	.00			.98

Total R^2 = .35, $P < .02$

Total Process Uniqueness = .24

Table 10

Commonality Results for Grade 3 Reading
(N=95 Classrooms)

			Commonalities			
	Construct	Uniqueness	12	16	26	Prob.
1.	Pretest	.09	-.04	-.03		.01
2.	Opportunity	.09	-.04		.03	.01
3.	Motivators	.02				.28
4.	Instructional Events	.01				.94
5.	Structure	.02				.79
6.	Setting	.08		-.03	.03	.002

Total R^2 = .29, $P < .01$

Total Process Uniqueness = .14

Table 11

Commonality Results for Grade 3 Mathematics
(N=119 Classrooms)

Construct	Uniqueness	Commonalities 25	126	235	Prob.
1. Pretest	.03		-.02		.12
2. Opportunity	.13	.02	-.02	.02	.0004
3. Motivators	.00			.02	.99
4. Instructional Events	.01				.83
5. Structure	.04	.02		.02	.39
6. Setting	.02		-.02		.07

Total R^2 = .30, $P < .003$

Total Process Uniqueness = .24

Table 12

Regression Coefficients and Correlations With Gain for Each Predictor Variable (Grade 1 Reading)

Variable	b	r_{ic}
1. Pretest: Mean	-.39	-.13
2. Pretest: Standard Deviation	.16	.03
3. Opportunity: Time	.20	.39
4. Opportunity: Curriculum	.43	.40
5. Motivators: Curriculum	.03	.15
6. Motivators: Interpersonal	.12	.16
7. Instruc. Events: Management	.05	-.05
8. Instruc. Events: Individual Cog.	.11	.10
9. Instruc. Events: Small Group Cog.	.00	-.03
10. Instruc. Events: Large Group Cog.	.01	.07
11. Instruc. Events: Quality	.01	.05
12. Structure: Placement	-.10	-.10
13. Structure: Mastery	.11	.11
14. Structure: Matching	.00	-.03
15. Structure: Assignment	-.06	-.01
16. Structure: Special Needs	.01	.15
17. Structure: Sequential & Pacing	-.02	-.01
18. Setting	.01	.09

$R^2 = .36$

$P < .001$

$N = 104$

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Table 13

Regression Coefficients and Correlations With Gain for Each Predictor Variable (Grade 1 Math)

Variable	b	r_{ic}
1. Pretest: Mean	-.41	-.18
2. Pretest: Standard Deviation	.27	.13
3. Opportunity: Time	.11	.11
4. Opportunity: Curriculum	.38	.25
5. Motivators: Curriculum	.17	.15
6. Motivators: Interpersonal	-.09	.05
7. Instruc. Events: Management	.16	.10
8. Instruc. Events: Individual Cog.	.16	.07
9. Instruc. Events: Small Group Cog.	.18	-.03
10. Instruc. Events: Large Group Cog.	.18	.12
11. Instruc. Events: Quality	-.06	.01
12. Structure: Placement	.10	-.00
13. Structure: Mastery	.07	-.04
14. Structure: Matching	.26	.25
15. Structure: Assignment	.04	.00
16. Structure: Special Needs	.01	-.07
17. Structure: Sequential & Pacing	-.09	.00
18. Setting	-.00	.05

$R^2 = .35$

$P < .02$

$N = 86$

Table 14

Regression Coefficients and Correlations With Gain for Each Predictor Variable (Grade 3 Reading)

Variable	b	r_{ic}
1. Pretest: Mean	-.22	-.12
2. Pretest: Standard Deviation	-.16	-.16
3. Opportunity: Time	.07	.17
4. Opportunity: Curriculum	.33	.29
5. Motivators: Curriculum	.16	.17
6. Motivators: Interpersonal	-.00	.05
7. Instruc. Events: Management	-.06	-.12
8. Instruc. Events: Individual Cog.	.05	.00
9. Instruc. Events: Small Group Cog.	.07	.01
10. Instruc. Events: Large Group Cog.	.07	.03
11. Instruc. Events: Quality	-.08	.10
12. Structure: Placement	.01	.03
13. Structure: Mastery	.10	.15
14. Structure: Matching	.01	.02
15. Structure: Assignment	.16	-.00
16. Structure: Special Needs	-.04	.06
17. Structure: Sequential & Pacing	-.03	.07
18. Setting	-.35	-.25

$R^2 = .29$

$P < .01$

$N = 109$

Table 15

Regression Coefficients and Correlations With Gain for Each Predictor Variable (Grade 3 Math)

Variable	b	r_{ic}
1. Pretest: Mean	-.27	.04
2. Pretest: Standard Deviation	.19	.12
3. Opportunity: Time	.12	.13
4. Opportunity Curriculum	.46	.38
5. Motivators: Curriculum	.00	.13
6. Motivators: Interpersonal	.01	.12
7. Instruc. Events: Management	-.05	.00
8. Instruc. Events: Individual Cog.	-.11	-.04
9. Instruc. Events: Small Group Cog.	.10	.09
10. Instruc. Events: Large Group Cog.	.09	.01
11. Instruc. Events: Quality	.00	.15
12. Structure: Placement	-.01	.03
13. Structure: Mastery	.08	.10
14. Structure: Matching	.17	.26
15. Structure: Assignment	.09	.02
16. Structure: Special Needs	-.05	.12
17. Structure: Sequential & Pacing	.11	.15
18. Setting	.17	.15

$R^2 = .30$

$P < .003$

$N = 119$

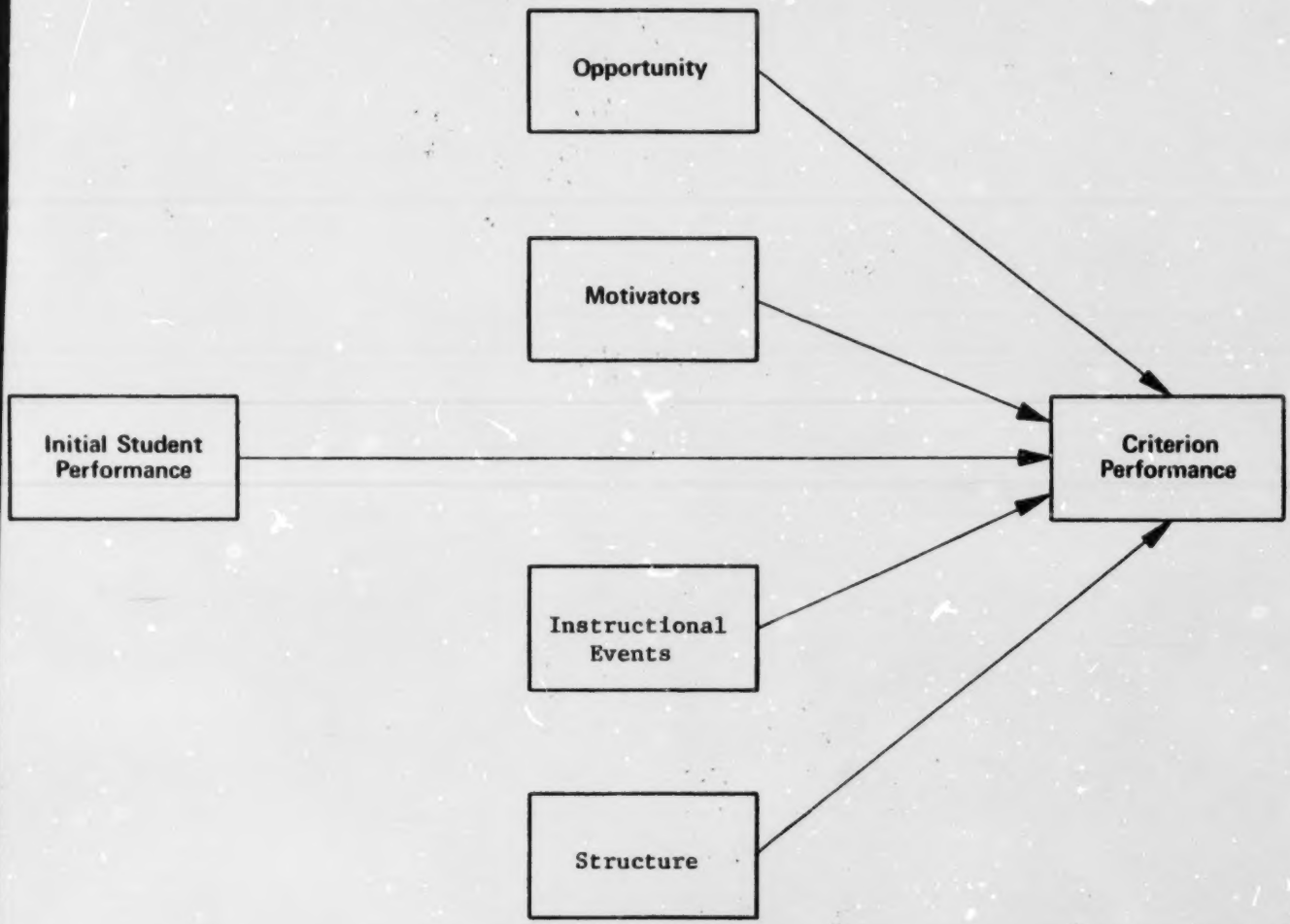

Figure 1
Model of Classroom Processes

Variable	#	Measure	Teacher Interview	Curriculum Analysis	Video Tape	Other
Time	1	number of students enrolled (-)	x			
	2	percent of students present	x			
	5	number of minutes per week in subj.	x			
	7	number of students transfer in (-)	x			
	8	number of students transfer out (-)	x			
	238	regular Instruction time				Master analysis file
	12	percent of students on task			x	
	239	days present				Expanded pupil roster
	240	days absent (-)				Expanded pupil roster
Curriculum	13	overlap of test with curriculum				Teacher estimate
	14	overlap of test with curriculum		x		

Figure 2

Map of Measures: Opportunity

Variable	#	Measure	Teacher Interview	Curriculum Analysis	Video Tape	Other
Curriculum	15	interest in curriculum material	x			
	22	number of modes of instruction - A		x		
	24	speed of correction	x			
Interpersonal	25	peer tutoring	x			
	26	use of games/contests	x			
	27	self-evaluation	x			
	28	self-management-seating	x			
	29	self-management-work alone	x			
	30	teacher contingent praise			x	
	31	teacher general praise			x	

Figure 3

Map of Measures: Motivators

Variable	#	Measure	Teacher Interview	Curriculum Analysis	Video tape	Other
Management	34	individual management statements			x	
	35	group management statements			x	
	36	individual cognitive management statements			x	
	37	group cognitive management statements			x	
Individual	38	individual cognitive statements			x	
	40	individual cognitive questions			x	
	43	individual child responses			x	
	45	ratio cognitive to management			x	
	46	individual cognitive contacts			x	
	54	extended tutor time			x	
Small Group	39	small group cognitive statements			x	
	41	small group cognitive questions			x	
	44	small group child responses			x	
	47	small group cognitive contacts			x	
	52	personal/tangential information			x	
Whole Class	48	whole class - cognitive statements			x	
	49	whole class - cognitive questions			x	
	50	whole class - child responses			x	
	55	lecture time (-)			x	
Quality	51	silence/non-verbal time (-)			x	
	59	active responses sought			x	
	61	refer to earlier curricular information			x	
	62	focuses child attention			x	
	63	solicit opinion on correctness			x	
	64	refer to earlier success			x	
	65	teacher tutors effectively			x	
	66	teacher manages effectively			x	
	67	child receives feedback			x	
	68	teacher interacts with individuals			x	

Figure 4

Map of Measures: Instructional Events

Variable	#	Measure	Teacher Interview	Curriculum Analysis	Video tape	Other
Placement & Pretest	69	curriculum-referenced placement test		x		
	70	placement procedures - standardized test	x			
	71	placement procedures - curriculum test	x			
	72	placement procedures - teacher test	x			
	74	placement procedures - observation of work sample	x			
	79	curriculum pretests		x		
Mastery	80	curriculum posttests		x		
	81	curriculum method assess mastery	x			
	84	teacher assesses mastery - curriculum test	x			
	85	teacher assesses mastery - teacher test	x			
	86	teacher assesses mastery - in-class assignment	x			
	88	teacher assesses mastery - completion of work units	x			
	89	teacher assesses mastery - observe work	x			
	91	teacher assesses mastery - conference with pupil	x			
Matching Practices	110	number of days since last test (-)				today's assignment
	111	range of days since last test (-)				today's assignment
	112	percent correct last test				today's assignment
	113	frequency of testing	x			
	116	tests used for assess mastery	x			
	117	tests used for diagnosis	x			
	118	tests used for prescription	x			
Grouping & Assignment	108	percent unique assignments				today's assignment
	109	to whom assignment made	x			
	134	percent time spent in entire class instruction (-)	x			
	135	percent time spent in subgroup of 11+ (-)	x			
	137	percent time spent in subgroup of 2-4	x			
	138	percent time spent in 1 to 1 instruction	x			
	141	number of students in groups (-)	x			
	142	frequency of regrouping	x			

Figure 5

Map of Measures: Structure

Variable	#	Measure	Teacher Interview	Curriculum Analysis	Video tape	Other
Meeting Individual Needs	94	when difficulty-curriculum suggests-additional practice		x		
	95	when difficulty-curriculum suggests-repeat incomplete		x		
	97	teacher alternative when difficulty - repeat material	x			
	101	teacher alternative when difficulty - 1 to 1 remedial	x			
	102	teacher alternative when difficulty - use curriculum suggested remedial	x			
	104	teacher alternative when difficulty - continue on, back later	x			
	145	readability level		x		
Sequence	121	clarity of sequence	x			
	123	teacher follows sequence	x			
	124	sequence varies within classroom	x			
	125	teacher create supplemental to improve sequence	x			
	126	each segment instruction prerequisite to next		x		
	127	different alternatives based on student's preference		x		
	130	self-pacing	x			
	132	rules/guidelines for pacing present		x		
	133	amount of review		x		

Figure 5 (continued)

Map of Measures: Structure